DOCUMENT RESUME

ED 040 945 24 SP 003 989

AUTHOR Koran, John J., Jr.
TITLE The Relative Effects of Imitation Versus Problem Solving on the Acquisition of Inquiry Behavior by Intern Teachers.
INSTITUTION Stanford Univ., Calif. Stanford Center for Research and Development in Teaching.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
REPORT NO TR-11
BUREAU NO BR-5-0252
PUB DATE May 70
CONTRACT OEC-6-10-078
NOTE 43p.

EDRS PRICE EDRS Price MF-$0.25 HC-$2.25
DESCRIPTORS *Behavior Change, *Instructional Films, Microteaching, *Problem Solving, Student Teachers, *Teacher Behavior, Teaching Models, Teaching Techniques

ABSTRACT

This study was designed to observe the effect of two basic variations of a film mediated model on teacher trainees' acquiring the technique of asking questions that require thought before being answered; it attempted to discover whether the subjects could best learn this technique by watching it illustrated by a filmed model, or by engaging in actual problem solving. It was hypothesized that training conditions which provide the most information regarding inquiring questions and which require the fewest steps to induce this behavior, produce greater behavior change than those which provide less information and require more steps to induce criterion behavior; also that training by model imitation produces greater behavioral change as measured by teacher performance in microteaching situations than does problem solving training; and that problem solving training with fewer steps to solution produces greater behavior change than that with more steps. Seven treatment groups and two control groups were used, with a total of 118 subjects. The results supported a null hypothesis since the increased use of the desired behavior displayed by the treatment groups was not significantly different from the control group. Trainee behavior tended to change in the direction predicted by the information processing model used. Data suggest that training methods were differentially effective for trainees who scored high or low on the initial pretest. (Author/MBM)

ED040945

STANFORD CENTER
FOR RESEARCH AND DEVELOPMENT
IN TEACHING

Technical Report No. 11

THE RELATIVE EFFECTS OF IMITATION VERSUS PROBLEM SOLVING ON THE ACQUISITION OF INQUIRY BEHAVIOR BY INTERN TEACHERS

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

John J. Koran, Jr.

School of Education
Stanford University
Stanford, California

May 1970

Published by the Stanford Center for Research and Development in Teaching, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position, policy, or endorsement of the Office of Education. (Contract No. OE-6-10-078, Project No. 5-0252-0501.)

SP003989

Table of Contents

	Page
Introductory Statement	v
Abstract	vii
Introduction	1
Related Theory and Research	2
The Dependent Variable: Inquiry Behavior	7
Hypothesis	8
Method	8
Subjects	8
Treatments	8
Rater Reliability	12
Results	14
Data Analysis	14
Hypothesis Tests	21
Discussion	22
Analysis of Procedures	25
Implications for Further Research	28
Implications for Educational Practice	28
References	31
Appendix	33

Introductory Statement

The central mission of the Stanford Center for Research and Development in Teaching is to contribute to the improvement of teaching in American schools. Given the urgency of the times, technological developments, and advances in knowledge from the behavioral sciences about teaching and learning, the Center works on the assumption that a fundamental reformulation of the future role of the teacher will take place. The Center's mission is to specify as clearly, and on as empirical a basis as possible, the direction of that reformulation, to help shape it, to fashion and validate programs for training and retraining teachers in accordance with it, and to develop and test materials and procedures for use in these new training programs.

The Center is at work in three interrelated problem areas: (a) Heuristic Teaching, which aims at promoting self-motivated and sustained inquiry in students, emphasizes affective as well as cognitive processes, and places a high premium upon the uniqueness of each pupil, teacher, and learning situation; (b) The Environment for Teaching, which aims at making schools more flexible so that pupils, teachers, and learning materials can be brought together in ways that take account of their many differences; and (c) Teaching the Disadvantaged, which aims to determine whether more heuristically oriented teachers and more open kinds of schools can and should be developed to improve the education of those currently labeled as the poor and the disadvantaged.

The following paper, Technical Report No. 11, reflects the concern of the Heuristic Teaching program with the ways in which teachers acquire the skills that enable them to teach effectively. The study was made as part of the project on Training Studies.

Abstract

The purpose of this study was to observe the effect of two basic variations of a film-mediated model on teacher trainees' acquiring the technique of asking inquiring questions, i.e., questions that require thought before being answered. The investigator wished to ascertain whether the subjects could best learn this technique by observing a filmed model which illustrated it, or by engaging in problem solving which required subjects to induce the type of questions asked and the probable answers.

It was hypothesized that: Training conditions which provide the most information regarding inquiring questions, and which require the fewest steps to induce this behavior, produce greater behavior change than training conditions which provide less information and require more steps to induce criterion behavior. It was further hypothesized that training by model imitation produces greater behavioral change as measured by teacher performance in microteaching situations than does problem-solving training, and that problem-solving training with fewer steps to solution produces greater behavior change than that with more steps to solution. One hundred and eighteen subjects, drawn from both science and nonscience populations, were divided into seven treatment groups and two control groups. The goal for each subject in each treatment was to recognize, identify, and reproduce the inquiring types of questions and answers.

The results supported a null hypothesis since the increased use of the desired behavior displayed by the treatment groups was not significantly different from group to group or from the baseline control group. However, trainee behavior tended to change in the direction predicted by the information-processing model used. Data suggest that training methods were differentially effective for trainees who scored high or low on the initial pretest. This might indicate further exploration with studies of trainee aptitutes' interaction with treatment methods.

THE RELATIVE EFFECTS OF IMITATION VERSUS PROBLEM SOLVING ON THE ACQUISITION OF INQUIRY BEHAVIOR BY INTERN TEACHERS

John J. Koran, Jr.[1]
University of Texas, Austin

The purpose of this study was to examine the relative efficacy of observing two basic variations of a film-mediated model on the acquisition of a teacher questioning technique called inquiry behavior, i.e., asking questions which require thought before being answered. The experiment was designed to assess whether Ss would acquire this questioning technique more efficiently: (a) by observing a film-mediated model which illustrates the asking of these questions; or (b) by being placed in a problem-solving situation which required Ss to induce the type of question asked and the probable answer.

This study emerged from the general concern for a more effective means of helping beginning teachers to acquire various verbal and nonverbal behaviors. It represents an attempt to explore an alternative to the usual strategies for the transmission or modification of teaching skills using written or oral instructions combined with discrimination training (Harris, 1963; Swearingen, 1962). In this latter approach, teacher trainees receive a written or oral description of the desired responses and their possible sequences. They subsequently attempt to make these responses in the classroom and receive periodic feedback from a supervisor regarding their relative success. This strategy is costly and inefficient, and under classroom conditions it is difficult for a supervisor to focus on specific teacher behaviors.

An alternative is suggested by the results of studies of observational learning and its influence on personality development (Bandura & Walters, 1963). These studies may be used as the basis for a training model which can be combined with one based on problem-solving strategies to produce acquisition of the desired behavior. The

---

[1]The research reported here was carried out while the author was a Research Assistant at the Stanford Center for Research and Development in Teaching. The author is presently an assistant professor in the Science Education Center at the University of Texas, Austin.

resulting study tests the relative effects of using both student and teacher behavior as training models, and using them together, alone, or as positive and negative instances of the behavior to be acquired.

One assumption that these techniques test is that by using student or teacher behavior alone, or negative instances of these, a problem-solving situation is produced which broadens a trainee's perceptions of what the stimulus condition (teacher behavior) or the response condition (student behavior) might be, and also suggests to the trainee what it is not. This method, it is predicted, produces more variety in the learned behavior and greater responsiveness to student behavior than the rather narrow range of skills which might be acquired by imitating a model depicting both the teacher questions and student answers.

Related Theory and Research

Imitation learning. The use of modeling procedures as the most likely means of producing learning is well documented. Bandura and Walters (1963) point out that new social responses may be acquired, or the characteristics of existing response hierarchies may be considerably modified, as a function of observing the behavior of others and the consequences of their responses without the observer performing any responses himself or receiving any direct reinforcement during the acquisition (Bandura, 1962).

Evidence has been gathered to suggest that the amount of learning exhibited by the observer can be as great as that shown by the reinforced performer (McBrearty, Marston, & Kanfer, 1961). Bandura and McDonald (1963) found that modeling cues are more effective than operant conditioning procedures under some conditions and that a model alone was as effective as the combination of modeling and reinforcement for initial learning. Bandura, Ross, and Ross (1963) have also gathered evidence that indicates that film-mediated models are as effective in producing behavior change as live models.

McDonald and Allen (1967), in a series of studies done with teacher trainees and using modeling procedures, have found that reinforcement and discrimination training administered by the experimenter were effective methods of producing behavior changes in teachers (videotape playbacks of a teacher's performance were used while reinforcing). Experiments using videotape teacher models also suggest that visual models are more effective than either reinforcement or discrimination training alone. Berliner, McDonald, Allen, and Sobol (1967), working with a dependent variable called higher-order questions, found that positive models of teacher-student interactions produced greater behavior change than negative models used imitatively.

Sheffield and Maccoby (1961) and McDonald and Allen (1967) have demonstrated that observational learning is greatly assisted by increasing the distinctiveness of relevant modeling stimuli. In the latter studies, videotape playbacks were used to focus the subject's attention on the correct response and to require him to attend to salient cues that occur during the course of learning. The playback feature of the videotape greatly facilitates the experimenter's attempts to emphasize cues, since the tape can be stopped or replayed at any point.

Problem solving. Corman (1957) points out that a problem exists when habitual responses fail to lead to the attainment of a desired goal. A search for a new response must then be undertaken. In problem-solving situations the subjects must not only search for possible alternative courses of action but must also select from among these alternatives the ones that will most successfully remove the obstacles to the goal. The search then is for information that will give structure to the problem. Presumably, as the amount of information generated is increased, the necessity for search is reduced. If all relevant information for a problem were available and understood, a problem would no longer exist. When little or no information is supplied, the problem solver may fail to develop the primary information needed.

If left to their own devices, problem solvers often set up false assumptions which make for unnecessary restrictions that slow down or prevent solution. The saving of time is a cogent argument for some guidance during the problem-solving sequence. Similarly, the problem-solving process becomes more efficient when the subject's attention is called to certain structural features of the problem situation rather than being given a general statement underlying a related cluster of problems. Performance in problem solving appears to improve as the amount of information given as guidance about methods of solution or specific principles for solution is increased.

In the present study, in which the subjects were to learn to ask inquiring questions, problem-solving circumstances were believed to exist in the following four conditions:

a. Under conditions in which the trainee observed a filmed model which depicted a teacher asking fact-oriented questions and students responding to them.

b. Under conditions in which the trainee observed a filmed model which depicted a teacher asking fact-oriented questions with no visible student response.

c. Under conditions in which the trainee observed a filmed model which depicted students answering fact-oriented questions, but no visible teacher questioning (stimulus).

d. Under conditions in which the trainee observed a filmed model which depicted students answering inquiring questions, but no visible teacher asking the questions (stimulus).

All of the above problem situations had in common the fact that the filmed (perceptual) model depicted an interaction or a portion of an interaction which could not be imitated by the trainee in order to acquire the desired behavior. Instead, the trainee was required to induce the desired behavior from the type of information he was provided by the visual display.

Each of the above cases satisfied the requirement of a problem situation as described by McDonald (1965). First, there was a goal to be attained in that all trainees were advised that their objective was to acquire knowledge of inquiring questions and answers and to be able to formulate inquiring questions. The trainees did not know the means of attaining the goal. A trained supervisor was provided to ask standardized cue questions and to guide the intern to attend to each stimulus or response condition while also eliciting verbalization of the trainee's observations and their inferences. Under these conditions, it was expected that: failure would be reduced, repetition of inappropriate responses would be reduced, motivation would be provided by the supervisor, and set would be established early in the training procedure by the initial instructions.

What evidence is available to support the above problem-solving model as a useful one for training? Gagné (1962) points out that requiring individuals to verbalize while in a problem-solving situation is significantly related to superior performance on the problem. Wittrock, Keislar, and Stern (1964) confirm this finding. Bandura (1966) describes similar observations from pitting viewers of a perceptual model who verbalized modeling stimuli against those who did not verbalize them, using reproducibility of the model's behavior as a criterion. He found that active symbolizers surpassed passive and non-symbolizers in reproducing more matching responses. His study provides supporting evidence for the facilitating role of symbolization on observational learning. McNeil (1965) also supports the role of overt responses and confirmation as a result of his studies in arithmetic. He found that overt responses and confirmation in the learning of mathematics problems were more effective than prompting procedures.

Use of positive and negative examples in problem solving. As Schultz (1960) points out, problem solving is high on the dimension of discovery of, as opposed to being told or shown, the correct response. This distinguishes problem solving from conditioning and rote learning where there is a minimum of response discovery. Unfortunately, there is precious little known of the effectiveness of problem-solving instructional methods vs. other instructional methods designed to achieve the same terminal behaviors. Duncan (1959) identifies as the most outstanding feature of the research in this area its failure to provide an articulate body of empirical relations. However, one area where there has been considerable research thought to be relevant to the study being described, is the use of positive and negative instances in concept formation. In this study the use of positive and negative models is thought to be analogous to the use of positive and negative instances in concept formation.

Smoke (1933) initiated the research on problem solving in the attainment of concepts with his study of the effects of providing instances of what the concept is not, along with, or as opposed to, the effects of information regarding what the concept is. Positive instances are usually defined as examples of the concept that include its essential characteristic, whereas negative instances lack one or more essential characteristic. The relative effectiveness of each for learning a concept is interpreted in terms of the number of steps required in each case before the trainee reaches the objective as inferred from the amount of information communicated by each instance (Hovland, 1952).

This principle was directly applicable to the design of training conditions for this study. It was possible, though, that with inquiry behavior as the dependent variable, those training conditions requiring the most steps or communicating the least information might have proven useful in a problem-solving training situation. It might have been that a single instance of negative information combined with induction and verbalization served to clarify the behavior in such a way that the trainee was able to reproduce a greater variety of behaviors than those training conditions with fewer steps to solution. In addition, those conditions which communicated the least information might in the long run have provided the most practice.

Hovland and Weiss (1953), Bruner, et. al, (1962), Glanzer and Huttenlocher (1960), and Huttenlocher (1962) have indeed demonstrated that the amount of information communicated per instance is important to the ease of problem solution. With conjunctive concepts, problems defined by all positive instances were more easily solved than those defined by all negative instances. Problems defined by a mixture of positive and negative instances were intermediate in difficulty. With disjunctive concepts, the information value of positive and negative instances reverses, negative instances generally transmitting more

information per instance. Here negative instances result in more efficient problem solving. The latter case seems to favor the training approaches used in this study.

Hovland (1952) proposes the following tools for evaluating the potential usefulness of different instances: If a problem involves an identifiable number of dimensions which may be relevant to the solution, and if each dimension can assume a limited number of values, one can analyze the potential value of information presented by the number of alternatives eliminated by the presented information. Each positive or negative instance presented to subjects eliminates some of the other possibilities.

Extrapolating from this, the effectiveness of a visual model could be measured by the number of elements eliminated by its presentation, or more simply, the amount of information communicated by this presentation.

Theory. The literature suggests that certain behaviors can indeed be learned by observing a model performing the desired behavior. Inquiry behavior presumably can be learned in the same way. The training method may become more powerful if verbalization is provided in the form of cue discrimination and feedback. But it is not known whether imitation or problem-solving conditions are more effective in producing acquisition of this complex verbal skill or what the relative effects are of the different types of information communicated.

The ultimate criterion for the evaluation of teacher performance has to be student performance (Medley & Metzel, 1963). Accordingly, one outcome of teacher training should be the senstivity of trainees to the consequences of the types of questions they ask, along with the ability to ask the type of question the training emphasizes. It is not certain that successful imitation of a teacher performing a complex task has alerted the trainee to which class of responses his questions elicited. It is possible that the range of behavior learned by subjects will not exceed that modeled. Observing either the desirable terminal behaviors or negative instances of teacher or pupil behavior does not insure that the trainee will induce and incorporate the behaviors necessary to elicit the desired student behavior. However, in the latter case, the problem-solving situation initiated may motivate the trainee to devise and to try responses which will elicit the desired kind of student behavior. Knowing what he is to produce or what he is not to produce may cause the trainee to try a broader range of questions, be more alert to when the behavior is or is not being elicited, and generally be more responsive and more adaptive.

The Dependent Variable: Inquiry Behavior

The dependent variable in this study is called inquiry behavior. It is defined for this purpose as a basic questioning technique which requires students to engage in complex cognitive processes. It is designed to appear in lessons where pupil participation is prerequisite to the goals of instruction. It is intended to make this participation a step to the attainment of concepts and principles and to engage students in inductive-deductive and problem-solving behavior.

The dependent variable was selected on the basis of work done by Gagné (1965) in the AAAS elementary school science curriculum. In these materials the inquiry behaviors are described as "basic processes" which enable one to acquire and process knowledge autonomously. These processes were then used to form classes of questions on all of the training materials.

Inquiring questions (variable 1). The teacher asks questions (elicits responses from students) which could be characterized as inquiry questions on the following list. The goal is to get students to think and respond in terms of the categories on the list. The 11 behaviors which characterize inquiring questions are:

a. Observational questions.
b. Classification questions.
c. Inferential questions.
d. Prediction questions.
e. Communication of ideas questions.
f. Theory or model questions.
g. Hypothesis questions.
h. Operational definition questions.
i. Manipulation and control of variables questions.
j. Planning an experiment questions.
k. Asking interpretation questions.

The teacher is attempting to guide students to employ critical thinking and inquiring behavior by asking the specific kinds of questions listed. The teacher's cues are the pupil's elicited responses and the verbal conditions immediately preceding these responses. This questioning behavior is characterized by asking for answers which are not available to the student or teacher through recall.

Fact-oriented questions (variable 2). Questions that elicit response from students which can be characterized as statements of facts or relating facts are called fact-oriented or noninquiring questions. These questions seek answers

that are known to both the student and the teacher, are convergent in nature, and are thought to be opposite in their effect on students from those called inquiring questions. They do not seek inductive, deductive, or divergent thinking as do inquiry questions.

Hypothesis

The general hypothesis tested was: Training conditions which provide the most information regarding inquiring questions, and which require the fewest steps to induce this behavior, produce greater behavior change than training conditions which provide less information and require more steps to induce the criterion behavior. It was further hypothesized that (a) training by model imitation produces greater behavioral change as measured by teacher performance in microteaching situations than does problem-solving training, and that (b) problem-solving training with fewer steps to solution produces greater behavior change than that with more steps to solution.

Method

Subjects

Subjects were drawn from a Stanford, University of California, San Jose State, and San Francisco State pool of secondary teacher trainees. They were initially stratified according to subject matter, and then randomly assigned to seven treatment and two control groups. The non-Stanford sample consisted of science teachers who were included in order to ascertain whether they were more or less receptive to training which emphasized a skill which was basically scientific in nature. All subject areas of the school curriculum were represented except foreign-language teachers since their audiotapes would have presented undue measurement problems. In addition, foreign-language teachers rarely use problem situations. The distribution of subjects by sex and subject matter is shown in Table 1. The total sample (N) was 128 which eventually was reduced by attrition to 118.

Treatments

As summarized in Table 2, the goal for each subject in each treatment was to recognize and identify the positive (inquiring) types of question and answer. The method of analysis was to note how much information each group was given and how much it required to reach the goal. The basic differences between Groups I, VI, and VII were that Groups I and VII provide all the information required to induce the positive behaviors. Subjects in Group VI were required to take two additional steps: (a) induce the positive questioning behavior from examples of the negative (fact-oriented) behavior and its consequences, and

TABLE 1

Distribution of Sample Subjects by Sex and Subject Matter

N = 128

Treatment Group	N	Sex M	Sex F	Combination	Subject Matter-Major[a] Eng.	Soc. Sci.	Math	Art Music	Stanford Calif. Science	State Colleges Science
I	15	5	10	1	4	4	1	1	2	2
II	14	6	8	1	4	3	1	1	2	2
III	14	9	5	1	4	3	1	1	2	2
IV	14	5	9	1	4	3	1	1	2	2
V	14	5	9	1	4	3	1	1	2	2
VI	14	4	10	1	4	3	1	1	2	2
VII	14	3	11	1	4	3	1	1	2	2
VIII	15	5	10	1	4	4	1	1	2	2
IX	14	4	10	1	4	3	1	1	2	2

[a]Drama subjects were grouped with English, and physical education subjects were distributed throughout the sample. In the breakdown they were included in the "combination." In practice, the combination group of subjects was made up of all subject areas.

(b) induce the types of answers that might be expected from the positive questions. Group VII members differed from Group I in that they read the essentials of the behavior, but subjects neither saw how the behavior should be performed nor what the consequences of the behavior would be. This presentation lacked a behavioral description of either the desired teacher behavior or its consequences.

Groups III and IV saw what the consequences of two types of questioning styles would be but were asked to induce the questions asked. Group III only required one step to reach the goal whereas Group IV required three additional steps to reach the goal. These included inducing: (a) the negative question; (b) the positive question; and (c) the positive answers.

TABLE 2

Amount and Type of Information Communicated

Treatment Group and Type of Model	Amount of Information Communicated	Amount of Information Needed
I S+ R+[a]	Two Steps[b]	Zero Steps[b]
II S+	One Step	One Step
III R+	One Step	One Step
IV R-	One Step	Three Steps
V S-	One Step	Three Steps
VI S- R-	Two Steps	Two Steps
VII S+ R+	Two Steps	Zero Steps

[a](S+) represents a positive teacher question; (R+) a positive student response; (S-) represents a negative teacher question; (R-) a negative student response.

[b]The steps recorded represent the amount of information given and amount needed to reach the goal.

Groups II and V only saw a teacher asking questions, but no evidence of the consequences of the questions. The difference between them is that Group II observers could imitate the types of questions asked and needed only to induce the correct response. This required one step. In Group V, three additional steps were required which included inducing the negative student behavior and from that the positive and negative behaviors. It is obvious that the amount and type of information communicated in each case varied, and patterns of information processing were suggested which were more or less demanding and perhaps of more or less value as training strategies.

As the previously discussed literature suggests, prior to this study, imitative models appeared to be an effective way to produce behavior change. The usual theory to explain this effect was contiguity theory. According to this explanation, subjects viewing a perceptual model of stimulus-response behaviors will learn the behaviors in sequence as they are presented due to the organization of the distinct elements in the task and without supporting reinforcement contingencies. The important factor here is the contiguity of the S-R. In this experiment contiguity theory may well explain the mechanism of Treatment I.

However, much more complicated processes are suggested by the other treatments as indicated by Table 1 and the previous theoretical discussion.

Prior to participation in the experiment all subjects were given instructions to prepare a problem-solving lesson which permitted them to ask questions. They arrived for training with this lesson and participated in the study on consecutive Saturdays. Table 3 describes the steps which followed on each treatment group and the amount of time for each stage of the training sequence.

TABLE 3

Summary of Steps in Treatment by Experimental Group[a]

Phase and Step in Treatment	Experimental Group									Minutes in Treatments
	I	II	III	IV	V	VI	VII	VIII	IX	
1. Introduction General Instructions	X	X	X	X	X	X	X	X	X	10
2. Teach (Pretest)	X	X	X	X	X	X	X	X	0	30
3. Written Model	X	X	X	X	X	X	X	0	0	10
4. Film-Mediated Model	X	X	X	X	X	X	0	0	0	30
or										
Written Model							X_1	0	0	30
or										
Placebo								X_2	X_3	30
5. Replan-Rehearse	X	X	X	X	X	X	X	X	X	10
6. Teach (Posttest)	X	X	X	X	X	X	X	X	X	30

[a]Symbols: (X) indicates that the subject received this phase. (0) indicates that the subjects did not receive this phase. The space under No. 4 indicates the difference between the film-mediated modeling treatment (X), the written model treatment (X_1), and the placebo treatment (X_2) and (X_3).

In terms of general procedure, then, all subjects except the control groups (VIII, IX) received preliminary instructions, were pretested, received written models, were exposed to the appropriate training procedure, and following a planning session, taught for the second time. Subjects were randomly assigned

to rooms, supervisors, and student microteams. Each subject spent a total of two hours in the training program. The subjects taught the same lesson to a different group of randomly selected students each time. This permitted subjects to retain the same basic subject matter in each lesson, while attending to improvements in questioning style.

Rater Reliability

Prior to the analysis of the experimental tapes, raters were trained on 14 pilot tapes. A rater protocol was developed to guide them. Reliability scores were derived for Judge 1, 2, and 3 over all categories. An analysis of variance model described by Winer (1963) was used to get reliability of the mean score for the three judges on each category. In addition, the mean interrater correlation between any pair of judges was also computed. Table 4 identifies the specific categories which were rated and the reliability of the mean scores along with the intercorrelation between raters for each teaching session.

Training procedures were planned to be evaluated in terms of their relative effects in increasing the frequency of asking inquiry questions. Table 4 shows the rater reliability for the total behavior and the subbehaviors which comprise it. For the most part, the reliability on the group category was extremely high and provided an effective measure for testing treatment effects. Since some of the categories showed lower reliability in a number of cases, it was decided to test the hypothesis using the total inquiry category rather than subsets of it.

Variation in reliability by category was due to a number of factors, but primary among these was the infrequency of particular classes of questions both before and after training. When manipulation of variables questions did occur, they frequently were considered prediction questions (i.e., what would happen if?) and were rated as predictions. Similarly, the distinction between operational definitions and inferences appeared to lie in the number of elements comprising the desired response. Hence, in most cases operational definitions and inferences were grouped together. All such incidences were rated as inference questions, since the students were required to infer the definition from the content of the lesson. The result was that the inference category, though highly reliable, probably contains many operational definitions. Most of these problems were anticipated before the study, and the most optimistic hope was that the training of raters would be effective enough to distinguish reliably between inquiry and fact-finding questions; this hope was realized.

In the fact-finding category, rote-and-recall answers were required. Changes in this category were thought to be another index of the training effect, for it was anticipated that an increase in inquiry questions would stimulate cognitive activity on the part of students and would result in lengthy responses which would cause a

TABLE 4

Rater Reliability

Category of Questions	Reliability of Mean Scores		Mean Interrater Correlation	
	Session I	Session II	Session I	Session II
Inquiring Total	.99	.99	.95	.95
Observation	.82	.96	.61	.89
Classification	.93	.96	.82	.86
Inference	.89	.91	.72	.77
Prediction	.84	.93	.69	.84
Communication	.75	.82	.64	.69
Models, Theories	.73	.56	.50	.33
Hypothesis	.99	.98	.65	.93
Operational Definitions	.00	.39	.09	.20
Manipulating Variables	.56	.83	.37	.75
Experimental Development	.95	.97	.88	.92
Interpreting Data	.99	.89	.77	.82
Noninquiring	.99	.96	.97	.94
Unclassified	.79	.78	.49	.55

decrease in the time for fact-finding questions. Furthermore, the training procedures were thought to be of a nature which highlighted what questions to ask and what questions to minimize. The total training effect was thought to be noticeable by an increase in variable 1 (inquiry questions) and a decrease in variable 2 (fact-finding questions).

Similar reasoning suggested that the ratio of inquiry to total questions (inquiry + fact finding) might better reflect the total training effect since this measure would be sensitive to, and reflect, simultaneous movements of inquiry and fact-finding questions.

Results

Data Analysis

The initial test used to determine if there were significant treatment effects on the criterion variable was an analysis of covariance. In this analysis, pretest scores were used as the covariate. This analysis was selected rather than the analysis of variance on pre- and posttest data because the pretest scores, when unadjusted, came close to being significantly different before the treatment. Since the covariate was a significant one, it was felt that this would provide a more "legitimate" although very rigorous test of the hypothesis. Here the experimenter chose to make a type II error.

The results of the analysis of covariance to test the major hypothesis are reported in Table 5. They indicate that there are no significant differences among the groups on the frequency of inquiry questions.

TABLE 5

Analyses of Covariance on Total Inquiring Questions Over Groups I to VIII

N = 104

Source of Variance	Sum of Squares	df	Mean Square	F-Ratio
Mean	243727.82	1	243727.82	
Covariate	15737.81	1	15737.81	
Treatment	1166.12	7	166.59	.93
Error	17040.00	95	179.36	
Total	277670.75			

As shown in Table 6, Treatment Groups I, III, IV, and VI show a definite increase in frequency from Session 1 to Session 2. Treatment Groups II, V, VII, and VIII started lower in the first session and failed to gain appreciably from Session 1 to Session 2. Treatment VIII, the no-treatment control group designed to test the practice effect of having taught a first lesson on the second lesson, indicates that no appreciable effect exists. Table 6 represents the result of t-tests

run on the differences from Session 1 to Session 2 to determine how significant the changes in frequency of the inquiring questioning behaviors were.

TABLE 6

Summary of the t-Tests of the Differences Between Session 1 and Session 2: Inquiring Questions

Treatment Group	Mean Scores		Standard Deviation		df	Critical Score	p
	S_1 [a]	S_2 [b]	S_1	S_2			
I	50	57	15	16	24	1.14	.12
II	39	38	14	17	22	.16	-
III	52	56	22	19	26	.53	.30
IV	46	50	11	20	24	.63	.25
V	41	42	15	16	24	.17	-
VI	48	55	17	20	24	.96	.20
VII	34	38	15	13	26	.78	.25
VIII	41	42	19	25	22	.12	-

N = 104

[a] S_1 = Session 1

[b] S_2 = Session 2

The critical scores come from students' t-distribution.

The table shows that none of the differences between Session 1 and Session 2 was significant at the .05 level. Treatments I and VI produced differences which were significant at the .12 and .20 levels, and Treatments III, IV, and VII also demonstrated gains which might be interpreted very liberally as being greater than chance. The treatments which produced the greatest changes were Treatments I, III, IV, and VII. This table must be interpreted in conjunction with the analysis reported in Table 5. It suggests that although there were no main effects which were significant between the groups, some of the treatment groups may have had stronger effects than others. Since the risk of a type II error is great with seven treatment groups and small N, these differences can only support further research with groups which seem to have a somewhat greater, though still weak, effect.

Analysis of covariance: Noninquiring questions. Table 7 presents the results of an analysis of covariance done on the noninquiring questions. The expectation here was that the treatment effects might have manifested themselves by a rise in total number of inquiring questions in some groups and a related decrease in total number of noninquiring questions. Table 7 shows that there were no significant differences between the treatment groups in the number of noninquiring questions asked.

TABLE 7

Analysis of Covariance on Total Noninquiring Questions Over Groups I to VIII

N = 104

Source of Variance	Sum of Squares	df	Mean Square	F-Ratio
Mean	85269.62	1	85269.62	
Covariate	6897.69	1	6897.69	
Treatment	1398.13	7	199.73	1.24
Error	15252.44	95	160.55	
Total	108817.88			

It is interesting to note in Table 8 that all of the treatment groups with the exception of Groups IV and VI, both using negative models, reduced the frequency of the undesired behavior, noninquiring questions. These two treatments appeared to produce a considerable rise in the frequency of the noninquiring questions from Session 1 to Session 2.

The significance of the change in frequency of the noninquiring questions using t-tests was also tested, noting again that this procedure was used only to see if in some groups differences might be occurring which are masked by the overall test. It is recognized that the hypotheses of the study cannot be tested by this procedure, nor can these results be treated as other than suggestive even when they are significant.

TABLE 8

Summary of the t-Tests of the Differences Between Session 1 and Session 2: Noninquiring Questions

Treatment Group	Mean Scores		Standard Deviation		df	Critical Score	p
	S_1[a]	S_2	S_1	S_2			
I	31	27	19	14	24	.63	.25
II	39	36	19	22	22	.36	-
III	36	34	23	16	26	.27	-
IV	22	28	13	13	24	1.17	.12
V	27	21	17	9	24	1.13	.12
VI	29	33	14	15	24	.79	.22
VII	24	23	13	10	26	.23	-
VIII	28	26	14	18	22	.30	-

N = 104

[a] S_1 and S_2 scores are mean scores for the session.

Although none of the differences in Table 8 were significant at the .05 level, it is interesting to note the pattern which is developing. Treatments I and VI show a change in the number of noninquiring questions which approaches significance. Previously these groups also showed an increase in the number of inquiring questions. Treatment I influenced the noninquiring questions in a negative direction, and Treatment VI moved them in a positive direction. A similar and slightly more pronounced change occured with Treatment IV and Treatment V; the former had an increase in inquiring questions which approached significance but also influenced the noninquiring questions in the same direction; the latter had little effect on the movement of inquiring questions but had an effect on the reduction in noninquiring questions.

Ratio of inquiring questions to total questions. Since several treatments appeared to influence the inquiring and noninquiring questioning frequencies in opposite directions, various ways to assess these changes were explored. One way was to examine the ratio of the inquiring questions to the total of inquiring and noninquiring questions. Table 9 shows the results of an analysis of covariance over all groups using the ratio score as the criterion score and the pretest scores as covariates. The table also indicates that the ratio of inquiring to total questions was not significantly different between treatment groups.

TABLE 9

Analysis of Covariance on the Ratio of Inquiring Questions to Total Questions

N = 104

Source of Variance	Sum of Squares	df	Mean Square	F-Ratio
Mean	419651.50	1	419651.50	
Covariate	6426.50	1	6426.50	
Treatment	596.19	7	85.17	.79
Error	10294.81	95	108.36	
Total	436969.00			

Table 10 shows that all of the treatment groups except IV and VII moved in the direction of increasing the inquiring questions and depressing the frequency of the noninquiring questions. Group VII did not show a great change from the previous figures; however, Group IV reflected the effect of both inquiring and noninquiring questions rising in frequency from Session 1 to Session 2.

Table 10 also assesses the relative degree of movement from Session 1 to Session 2 by using t-tests of the significance of the changes with the caveats previously described.

It is interesting to note in Table 10 that the imitation model shows an effect in increasing the desired behavior and decreasing the undesirable one. In addition, the treatment using the negative student model shows a comparable but opposite effect, while the other training methods show very weak effects.

Comparison of control Group IX with pretest scores. Another analysis compared the Treatment IX posttest only with the pretests in the other groups. The question was asked: Do Treatment IX scores significantly differ from the pretest scores? These control group scores were not meant to be compared with the posttest scores of the other groups because such a comparison would only serve the same purpose as comparing the pretests with the posttests in the analyses of covariance. Table 11 shows the results of comparing the post-test with the pretests of the other groups.

TABLE 10

Summary of t-Tests of the Differences, Session 1 to Session 2 on the Ratio of Inquiring to Total Questions

Treatment Group	Mean Scores S_1[a]	Mean Scores S_2	Standard Deviation S_1	Standard Deviation S_2	df	Critical Score	p
I	64	69	11	9	24	1.28	.10
II	42	54	17	19	22	.96	.15
III	62	63	15	15	26	.18	-
IV	70	65	12	12	24	.11	-
V	61	67	15	10	24	1.20	.11
VI	64	65	10	10	24	.26	-
VII	61	62	14	10	26	.22	-
VIII	58	61	14	19	22	.41	-

N = 104

[a] S_1 and S_2 scores are mean scores for the session.

TABLE 11

A Comparison of Group IX with Pretest Scores: Analysis of Variance

N = 118

Source of Variance	Sum of Squares	df	Mean Square	F-Ratio
Between Groups	3766.5586	8	470.82	1.80
Within Groups	28565.4648	109	262.07	
Total	32332.0234	117		

The F-ratio indicates that the Treatment IX control group scores did not differ significantly from the other pretest scores. This suggests that little happened to the Group IX people over the duration of the training sequence to influence their behavior. Group IX subjects finished the training sequence with scores that were not significantly different from those of the other groups in the pretest. The high F-ratio is understandable due to the random incidence of a very low group, Group II, in the pretests, which appeared to be different from the other groups initially.

Finally, an examination of the interaction between subject areas, divided into four groups comprising science and nonscience subjects and treatment, showed that there was no significant interaction present. In addition, the non-Stanford science sample appeared to respond to the training procedures in the same way as the Stanford sample. Table 12 presents a summary of the data from this analysis.

TABLE 12

Interactions Between Subject Taught and Treatment Effect

N = 104

Source of Variance	Sum of Squares	df	Mean Square	F-Ratio
Mean	22552.12	1	22552.12	
Covariate	237140.96	1	237140.96	
Treatment Effect	951.44	7	135.92	.76
Subject Effect	169.94	1	169.94	.91
Interaction	515.69	7	185.52	.98
Error	16340.60	87	187.67	
Total	277670.75			

Hypothesis Tests

The foregoing data indicate that the null hypothesis was not rejected in this study with regard to main treatment effects. None of the treatments was significantly different from the other or from the baseline control group.

A comparison of Group IX with the other pretests shows no significant difference at the .05 level. This means that during the training period there was no significant influence on the criterion behavior other than that of training. Similarly, the practice effect of having taught the first lesson with the criterion behavior in mind did not seem to influence performance on the second lesson when no treatment intervened (Group VIII).

An examination of the tables shows that on the inquiring variable all treatments but II (teacher positive model only) increased from pretest to posttest. Only Treatments I, III, IV, VI, and VII had effects which approached a significant level. Table 13 shows the treatments with the most effects on the criterion variable, inquiring questions.

TABLE 13

Summary of Training Effects on Inquiring Questions
Session 1 to Session 2

Type of Model		Relative Effect p
I	positive combination	.12
III	positive students only	.30
IV	negative students only	.25
VI	negative combination	.20
VII	written model	.25

Although noninquiring questions went down in all groups, two groups showed unexpected results. Groups IV and VI, using the negative students-only model and the negative combination model, each increased in the frequency of the undesired type of questions (factual). Treatment I probably produced some effect in reducing the number of noninquiring questions ($p < .25$) along

with the effect pictured in Table 13 on the inquiring questions. It is interesting to note that Treatment V, which had no noticeable effect on increasing the inquiring question frequency, probably decreased the number of noninquiring questions ($p < .12$).

An examination of the ratios of inquiring questions to total questions showed that Treatments I, II, and V might have some effect. The other treatments showed negligible effects in their total influence.

Discussion

This study sought to determine the effects of various types of presentation variables emphasizing modeling and problem-solving conditions on the acquisition and increase in the frequency of inquiry questions. The general hypothesis tested was: Training conditions which provide the most information regarding inquiring questions, and which require the fewest steps to induce this behavior, produce greater change than training conditions which provide less information and require more steps to induce the criterion behavior. It was further hypothesized that (a) training by model imitation produces greater behavioral change as measured by teacher performance in microteaching situations than does problem-solving training, and (b) problem-solving training with fewer steps to solution produces greater behavior change than that with more steps to solution.

The general assumption from which the hypothesis was derived was that the rate and level of learning of a given teacher strategy varies as a function of model presentation. Specific predictions, though tentative, were based on theoretical considerations which suggested that the differential effectiveness of various model types might stem from the amount and type of information communicated by each model. Training procedures varied along a continuum from positive stimulus-response to negative stimulus-response questioning conditions. Variations in the mode of presentation involved observing only a stimulus or reponse of the positive and negative type, or observations of a combination of positive instances and one of negative instance, to only a written model.

The results of the data analyses are more suggestive than conclusive. However, in those cases in which the subjects were trained by imitating a model, either written or film mediated, there were slight, but not significant, upward movements. Problem-solving models requiring the least information to be given (Group VI) also made impressive but not significant gains in the frequency of the criterion behavior. The former of these phenomena provides some evidence of the efficacy of imitation as a training procedure. In addition, however, the latter results suggest that problem-solving, under some conditions and with certain dependent variables similar to those used, may have some training potential. Further research is necessary in both cases.

The fact that the experimenter could not reject the null hypothesis indicates that the hypothesized effects, although they occurred to some extent, did not occur to a significant extent. Table 6 indicates the direction and the magnitude of the changes. Measures of the extent of these changes from Session 1 to Session 2 were made as a means of describing the data only. It is recognized that change scores of this type are not an extremely reliable measure, but important information regarding the effects of the individual treatments can be inferred through a cautious use of these scores.

Treatment Group I was thought to be a maximum information group and tended to have the strongest effect in increasing the frequency of inquiry questions and decreasing the frequency of noninquiry questions. This was an imitation group which showed the teacher asking desirable questions and the students making desirable responses. In this group the stimuli and the responses were contiguous. The types of questions that should be asked were followed by the kind of response that could be expected if these questions were asked. Possibly a matching response was also present in which the subject compared the responses the model was getting and the stimuli that were producing them, with the responses he elicited on the pretest. It is also possible that viewing student responses which are desirable and similar to some of the responses the subject elicited during his first teach were reinforcing, causing positive behavior to be repeated in teach two. It seems that observing the correct sequence of teacher-student behavior also serves to provide valuable information which aids in decreasing the occurrence of undesirable behavior (Table 8).

Treatment Groups II and III showed trends which appeared to be consistent with theory and also with the Treatment I effects. In Treatment II the same teacher was to be imitated as in Treatment I and in Treatment III the same student responses were shown as in I. The weak changes which occurred in these treatments were congruent with expectations. The teacher-only model (Treatment II) did not produce a change in desirable teacher behavior from Session 1 to Session 2, but it did produce a weak decrease in the frequency of the noninquiry questions. It may be that the behavioral conception presented here was insufficient for imitation purposes but adequate for matching purposes. Hence, the reduction of noninquiry questions.

The student-only Treatment III, on the other hand, showed a weak effect on increasing inquiry questions ($p < .30$) and a weaker effect in changing the noninquiry questions. This might indicate that viewing the student and teacher contiguously combined the relative effect of viewing the students alone in increasing the inquiry questions with the combined effects of the teacher and student models alone in decreasing the frequency of noninquiry questions. The mechanisms here appeared to include:

a. Contiguity of stimulus-response in Treatment I and a complete behavioral conception which provided maximum information.

b. Student feedback acting as a reinforcer to the observer when he compared these responses to the responses he received during his teaching Session 1.

c. A lower amount of information communicated by Treatment II and III models than in I.

Table 10 suggests that the matching behavior which may have occurred when observers saw only a teacher performing the desirable behavior seemed to influence the ratio of inquiry to total scores. Although the teacher-only group (II) had little effect in producing the desired behavior, its total effect was relatively stronger ($p < .15$). It appears necessary from this to have the teacher-student behaviors presented contiguously since seeing the teacher alone only appears to serve as a standard with which to compare one's own previous questions and to induce the behaviors that should not be imitated. It may also be possible that verbalization of the induced stimulus-and-response conditions in both Treatments II and III evidently did not provide the contiguity that a visual display of both behaviors does.

Treatments IV, V, and VI, all negative presentations thought to convey little information, suggested some interesting processes taking place. Treatments IV and V showed the student or teacher only whereas VI combined the students and teachers. In the first two there were three steps to identification of the desirable behavior, whereas the latter required two. Verbalization in Treatment IV included stating what kind of answer was observed, what kind of question produced it, and generation of how an inquiry question could be asked. Treatment VI received more information than any other negative group in that it presented a behavioral conception of what the incorrect questions and answers looked like. It was thought that this behavioral conception was contrasted with the subjects' teach one and functioned to assist the subject to induce the desired behavior for the second teach.

The surprising outcome was that both groups, IV and VI, increased the frequency of inquiry questions ($p < .25$ and $p < .20$), but at the same time increased the frequency of noninquiry question ($p < .12$ and $p < .22$). What seems to have happened was that in addition to these groups acting as was predicted from the information-processing model, they also acted as models of the undesirable behavior and were imitated. The elements common to these three treatments were the observer's verbalizations of what he saw and what it suggested to him about the skill to be acquired. It is possible that the visual-verbal contiguity established by visualizing and contrasting a response to the one observed, and then the stimulus that could produce it, functioned to produce the above effects.

Treatment V did not influence the inquiry questions but had an effect ($p < .12$) in decreasing the fact-oriented questions. As in the positive case, Group II, it may be that viewing the teacher alone served to set up a comparison situation whereby the subject could either be reinforced for learning what the right types of questions are (Treatment II) or use the information about what the wrong questions are to produce reduction in this class of behaviors. In both the positive and negative case, the teacher model alone seemed to be of dubious value. For imitation and problem solving to occur then, the teacher-student models that showed the behaviors contiguously and communicated the most information, appeared to be the more effective ones.

Treatment VII had some effect ($p < .25$) on increasing the inquiry questions but little effect on decreasing the negative questions. Evidently, the lack of a behavioral conception here did not seriously inhibit the ability of the subjects to recognize the way desirable questions should be asked. The written model did not seem to provide the kind of information necessary for the subject to understand what types of questions not to ask. Since no behavioral conception was present here, it can be inferred that the verbalization which occurred in this treatment rather than the visual matching with the subject's own previous teaching image was the influencing factor. This is useful information because it suggests that perhaps verbalization in the other models was a more influential process than the possible matching behavior. Although Treatment VII was an imitation condition it was thought to be minimal because it lacked the behavioral conception.

Treatment VIII indicated that the criterion behaviors were not influenced to a great extent by the practice effect of having taught the first lesson. Both inquiry and fact-oriented questions fluctuated one and two points respectively in a positive and negative direction. Treatment IX was compared with the pretests of the other treatments and found not to differ significantly. Both of these control groups provided evidence which suggested that experimental conditions other than the treatments had little influence on the frequency of the dependent variable.

Analysis of Procedures

Whenever nonsignificant differences result from an experimental study, it is incumbent upon the experimenter to pinpoint sources of variance. Much thought has been given to the present experiment and indeed to the entire theory of social learning as applied to research in teacher education. As a result, a number of generalizations have emerged which are categorized as follows: (a) Independent variable, (b) research design, and (c) dependent variable. Each of these will be considered within the frame of reference of research in teacher education.

Independent variable. There is considerable research both in areas of behavior modification and teacher education to suggest the efficacy of imitation learning. A common characteristic of many of these studies is an experimental design which pits a strong imitation treatment with a weak control. Although two weak controls were used in this study, the relative strength of the behaviors to be imitated did not exceed them. The nature of the behavior depicted by the model appeared to contribute to the weak effects of the models. The behavior to be imitated had 11 subbehaviors to be acquired over a period of 30 minutes. Many of these subbehaviors, as suggested by the reliability scores, were very much alike. Hence, a major problem here could well have been that some stimuli were more distinct than others. At the same time, the longer training time (30 minutes) may easily have resulted in a memory problem or a cognitive overload, i.e., behaviors coming earlier in the sequence may have been forgotten. Another problem contributing to the possible lack of power of the models was the fact that the behavior to be acquired, inquiry questioning, was difficult to highlight and distinguish from fact-oriented questioning in a meaningful model lesson. Hence, the positive and negative models may have been quite alike in terms of stimulus value.

The supervisor's role may also have been a source of interference. Although supervisors had scripts which told them what to do and say and were randomly assigned to subjects, it was impossible to monitor their behavior. Consequently, they may have caused any number of unaccountable effects which could have had a leveling influence on the effects of the models. It would have been wiser to place supervisor comments on the sound track of the videotape along with the model, and to restrict supervisor activities to guiding the subjects to their rooms and handing out and collecting materials.

Another source of influence could have been the set induction materials handed to subjects prior to the training. It is possible that the very nature of these descriptive materials served to cue and direct trainee attention to the behaviors in question prior to the pretest. If this did occur, it would account for the high frequency of the behavior to be acquired on the pretest and the subsequent leveling off of the behavior on the posttest.

Research design. This study had nine groups to which subjects were assigned randomly. A larger than usual number of treatment groups was employed because it was expected that the total number of subjects would be twice the sample that materialized. Had the sample size been known, or more accurately predicted, a smaller number of treatment groups would have been used. The significance of the number of treatment groups is manifold. Since teaching and learning experiments frequently have a large variance among subjects, the larger sample would have influenced the needed F-ratio for significance; at the same time, smaller samples have a greater probability of variance which cannot be explained. For this reason, studies of this type should always include

the use of covariates which can provide the researcher with an opportunity to explain more of the variance by adjusting the pretest scores for factors described by the covariate. In this study the small number of subjects per treatment, in combination with the large variance which was unaccounted for, increased the probability of the type II error.

Finally, studies of this kind require a series of trials, perhaps three, in order to get a clear picture of what is taking place in each trial and over time. Two sessions were planned because it was felt that a major question in this study was which treatment was most effective for the acquisition of the skill through imitation on the first exposure. The rationale here was that the movement from trial one to trial two in most imitation studies is usually the most dramatic. Beyond trial two, there is usually a decrease in the strength of the criterion behavior. However, when the dependent variable is a complex one, a number of trials may be necessary to produce the optimal effects.

Dependent variable. The dependent variable was a class of questions identified as inquiry questions. They were contrasted with fact-oriented or memory questions. There were actually 11 subclasses of questions comprising the former group and these were presented as questions, answers, or a combination of these, in the positive models. The negative models presented fact-oriented questions, answers, or a combination. A major difficulty with this dependent variable is its complexity. Over a 30-minute period, the observation of an inquiry model may well have produced a cognitive overload. In addition, the logical sequential nature of these questions may have imposed undue strain on trainees if they were unable to recall basic question categories, such as observation and classification, on which to build later questioning types, such as experimental development and interpreting data. The nature of this class of behaviors may have been such that it frustrated observers, bored them, or simply provided too much in too short a time. Future studies might do well to decrease the size of the dependent variable and make the categories as distinctive as possible.

Another interpretation is that the results derived in this study suggest that the behaviors modeled were already present to a great degree in the observers' response repertoire, hence the effects of training were negligible. For example, the introductory materials to trainees described the behaviors to be acquired so completely that the introduction became a training package in itself. This was discussed earlier.

In conclusion, a word must be said about rating procedures. Whenever behaviors of the type examined here are used in a study, it is essential to define the categories to be rated in a way which is specific to the way the behaviors were modeled. It may be that the behaviors rated in this study were broader than those modeled, in which case the effects of modeling could have been negated.

The previous discussion has been presented as speculation regarding the lack of more than suggestive results in this study. Many of the possible sources of unexplained variance in the data are influences which many researchers will, or have, encountered, and are provided here to present a record on which future research can be built. If such a record had been available prior to the conceptualization and planning of this study, the outcome might have been of a different nature. It is the investigator's firm conviction that in order for research to answer some of the important questions and criticisms leveled at educators, the reporting of theoretically broad studies, the data derived, and results of both conclusive and inconclusive studies must be encouraged and widely disseminated.

Implications for Further Research

The surface has barely been scratched in the area of using videotape technology in teacher training. There seems to be an abundant supply of teacher behaviors which could be used as dependent variables. Anything a teacher does or says which improves student learning is a potential skill to be learned by preservice and inservice teachers. In addition, the range of independent variables extends over the entire spectrum of the psychological variables which concern learning. Investigators need to look more closely at the nature of effective perceptual models and the effects of various types and amounts of set induction, practice, feedback, and reinforcement. Another area which needs examination is: To what extent are skills remembered and used after preprofessional training is over, and what are the effects of the subjects' value system and those of the schools they staff on subsequent performance of the skill?

This study has generated more questions than answers regarding some of these areas. For instance, how effective is the problem-solving method when used in the acquisition of certain skills? What is the effect of observing student behavior on subsequent teacher performance? Questioning behavior is an important instructional mode. It can be, and should be, experimented with, one dimension at a time, to determine the precise influence of certain training methods on the assumption of certain types of questions. Experimental designs can move toward fewer treatment groups and greater numbers of subjects in each when dealing with teacher trainees as subjects. Finally, an area on the horizon which appears to be a key to the solution of the variability problem within a teacher training sample is that of the interactions which may exist between a given presentation variable and a particular trainee aptitude. This area must surely be explored.

Implications for Educational Practice

Research cannot go on in a vacuum. The goal of research is application in the schools. This study employed a training method and a set of conditions

which could be used in the schools as a means of inservice training, colleague evaluation, curriculum sequencing and development, and supervision. It must be remembered, though, that the information derived from this study and accumulated from previously completed studies merely suggest these applications. It remains to be demonstrated which are the most effective means of inservice training, teacher evaluation, and supervision and then to devise means of introducing these findings to the schools. In education, as in industry, research must precede development, not the reverse.

References

Bandura, A., & McDonald, F. J. The influence of social reinforcement and the behavior of models in shaping children's moral judgments. Journal of Abnormal and Social Psychology, 1963, 67, 274-281.

Bandura, A. Observational learning as a function of symbolization and incentive set. Child Development, 1966, 37, 499-506.

Bandura, A., & Walters, R. H. Social learning and personality development. New York: Holt, Rinehart and Winston, 1963.

Bandura, A., Ross, D., & Ross, S. A. Transmission of aggression through imitation of aggressive models. Journal of Abnormal and Social Psychology, 1963, 66, 3-11.

Berliner, D. C., McDonald, F. J., Allen, D. W., & Sobol, F. T. A comparison of different modeling procedures in the acquisition of a teaching skill. Paper presented at the American Educational Research Association, New York, February, 1967.

Bruner, J. S., Goodnow, J., & Austin, G. A study of thinking. New York: Science Editions Inc., 1962.

Corman, B. R., The effects of varying amounts and kinds of information as guidance in problem solving. Psychological Monographs, 1957, 71 (2).

Duncan, C. P. Recent research on human problem solving. Psychological Bulletin, 1959, 56, 397-429.

Gagné, R. M., & Smith, E. C., Jr. A study of the effects of verbalization on problem solving. Journal of Experimental Psychology, 1962, 63, 12-18.

Gagné, R. M. The psychological basis of science: A process approach. Washington, D. C.: American Association for the Advancement of Science, 1965.

Glanzer, M., & Huttenlocher, J. The mechanics of concept-work, concept size, example sign, series complexity, and information order. Washington, D. C.: Office of the Surgeon General Technical Report, November, 1960.

Harris, B. M. Supervisory behavior in education. Englewood Cliffs, N. J.: Prentice Hall, Inc., 1963.

Hovland, C. I., A communication analysis of concept learning. Psychological Review, 1952, 59, 461-472.

Hovland, C. I., & Weiss, W. Transmission of information concerning concepts through positive and negative instances. Journal of Experimental Psychology, 1953, 45, 175-182.

Huttenlocher, J. Some effects of negative instances on the formation of simple concepts. Psychological Reports, 1962, 11, 35-42.

McBrearty, J. F., Marston, A. F., & Kanfer, F. H. Conditioning a verbal operant in a group setting: Direct versus vicarious reinforcement. American Psychologist, 1961, 16, 425-434.

McDonald, F. J. Educational psychology. Belmont, Calif.: Wadsworth Publishing Company, Inc., 1965.

McDonald, F. J., & Allen, D. W. Training effects of feedback and modeling procedures on teaching performance. Technical Report No. 3. Stanford Center for Research and Development in Teaching, 1967.

McNeil, J. D. Prompting versus intermittent confirmation in the learning of a mathematics task. Arithmetic Teacher, 1965, 12, 533-536 N.

Medley, D. M., & Mitzel, H. E. Measuring classroom behavior by systematic observation. In N. L. Gage (Ed.) Handbook of Research in Teaching. Chicago: Rand, McNally Company, 1963. Pp. 247-328.

Schulz, R.W. Problem solving behavior and transfer. Harvard Educational Review, 1960, 30, 61-77.

Sheffield, F. D., & Maccoby, M. Summary and interpretation of research on organizational principles in constructing filmed demonstrations. In A. A. Lumsdaine (Ed.) Student response in programmed instruction. Washington, D. C.: National Research Council, 1961. Pp. 117-140.

Smoke, K. L. Negative instances in concept learning. Journal of Experimental Psychology, 1933, 16, 583-588.

Swearingen, M. E. Supervision of instruction: Foundations and dimensions. Boston: Allyn and Bacon, Inc., 1962.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill Book Company, 1962.

APPENDIX I

Rater Protocol: Revised Rating Criteria

Experiment No. 10

GENERAL RULES

1. When a teacher is eliciting a response, he is asking a question.

2. If a question extends over more than one line, record according to the line on which the question ends.

3. Rate each inquiry as given. Don't add punctuation even if additional punctuation seems necessary to "make sense" out of the inquiry.

4. If inquiries have blanks or data indicating something has been omitted, rate anyway as long as the inquiry makes sense.

5. If cue words, i.e., "observation," "inference," "hypothesis," are used in a questioning context, rate the inquiry according to the key word.

6. Frame of reference determines whether or not an unpunctuated, possible inquiring statement should be rated. If the phrase is part of a series of inquiries, rate it; if it isn't, don't rate it. Examples of phrases that probably should be rated include, "another example," "another one," "another rule," "go ahead," "keep going," etc., even if they can be rated as yes or no questions.

7. Rate single inquiring words such as "how" or "why" as the category for which they are a cue.

8. Inquiries such as "What else?", "Can you give another example?" are considered to be extensions of a previous question and are rated as the same category as the question of which they are an extension even if they would be rated as yes-no questions.

9. Both clauses of questioning compound sentences separated by "and" are rated.

10. When there is a series of questioning words beginning an inquiry or there are inquiring words separated by "or" which cause the question to be rated in two or more different categories, classify on the basis of the last inquiring word; i.e., "How or why do that?"

11. Repeats are identical inquiries following an initial question. If words or meaning changes with the second inquiry or there is an intervening student talk, rate both the inquiry and the second question separately.

I. INQUIRY

The teacher asks questions of the students which can be categorized into one of the following eleven categories:

A. Induction

1. Systematic Observation: Observing relations under conditions in which one of more physical characteristics vary as detectable by any of the senses. ("Vary" refers to existing differences in the environment which the student is asked to observe, not to the human manipulation of variables.) Questions may encourage general observation or observation when conditions have changed.

   Includes:

   1. Description of the environment based on physical characteristics detectable by any of the senses past or present -- abstracting physical characteristics.
      a. Describe what you have in your hand.
      b. What do you see?
      c. Would you describe this object?
   2. Systematic observations in terms of standardized measurement units (e.g., length, width, volume, temp.) or graphs.
   3. Systematic observation in terms of identifying the effects of motion, time, space. Students may be asked to express observed conditions by means of verbal and written number sentences, and identify and recognize the effects of motion as applied to dimensions of effort, distance, direction, time, appearance.
      a. How could you express that difference in terms of time?
   4. If inqiuries beginning with "What about..." are trying to elicit an observation, rate as such.

B. Generalization

1. Systematic Classification: Teacher seeks answers from students devising classifications of objects by means of single or multiple dimensions such as state of matter, color, volume, symmetry, area, and weight.

Includes:

1. Placing objects or ideas into groups or categories.
2. Identifying whether objects or thoughts are the same or different.
3. Comparisons are classification.
4. Examples are classification.
5. Asking for similarities or differences is classification.
6. Characterizing objects or behavior.
7. Questions in which teacher gives alternatives or possible answers.

Examples:

1. According to what characteristics would you group these objects?
2. Into what categories would Hemingway's writing fall?
3. Are they the same?
4. How would you categorize them?
5. On what bases could you compare these?
6. Would you give an example of each type?
7. Could you state the similarities and differences between these objects?
8. Describe her for me.
9. Characterize them
10. Is that fact or inference? Will you raise or lower prices?

2. Inferential Statements: Teacher seeks responses from students drawing reasonable inferences about phenomena and distinguishing them from the observations on which they are based.

   Includes:

   1. Questions which call for a critique of a response or condition.
   2. Questions in which data is given and student told to apply it.
   3. Questions which elicit responses leading to the identification of a problem in nonexperimental circumstances.
   4. Think questions calling for interpretation in nonexperimental contexts.
   5. "Higher order definitions" (definitions not available to recall or those developed in the lesson causing student to think).
   6. "When" questions if observations or data to infer or comment on is given.
   7. "How" questions calling for a single statement or thought process.
   8. Questions beginning with "what about..." that try to elicit an inference.

9. In many instances, "should" or "could" questions unless another seems appropriate.
10. In many instances, questions beginning with "What about...," "What if...," "How about...," "So, what?"

Examples:

1. What do you think about that?
2. What do you need to know about her?
3. What do you think they do agree on?
4. What are your thoughts about the situation? What do you mean?
5. What is a good teacher?
6. When does the value go down?
7. How do I know this? How is she feeling?
8. What about her fear of elevators?
9. What do you think he could say?
10. What if nothing happens? How about raising the temperature?

3. Predictive Statements: Teacher seeks a prediction of a specific outcome or effect based upon an observed set of events.

Includes:

1. All questions in which events are given and the student is asked to predict an outcome, or prediction of outcome or consequences alone are asked for.
2. The future tense is frequently an indication that a prediction is being elicited.
3. In many instances, "would" questions, unless a key word for another category is given.
4. Questions in which no item is varied or changed and in which an outcome is to be given or assumed.

Examples:

1. What is going to happen?
2. From this experiment, what would you predict would happen when this substance is heated?
3. What do you think he will do next?
4. What is she going to do?
5. What would happen if she got into the elevator?

4. Communicating Ideas: Teacher seeks answers from students communicating a series of observations or ideas from one individual to another. Questions of this type call for an oral answer describing changes in physical state, motion, color, weight, volume, etc.

Includes:

1. Summaries given by student. Summary of lesson material.
2. Including or mentioning other students as recipients of summaries.
3. Any summary behavior. Student-student; student-teacher.
4. Restatements of previously discussed material.

Examples:

1. Could you summarize what we have been talking about to the class?
2. Would you describe to the rest of the class what we saw?
3. Tell me what we've said so far.
4. What have we said so far?

5. Models and Theories: Teacher seeks responses from students which combine observations, communications, and inferences into a theory or model, or in which students state a generalization holding true over a number of cases.

Includes:

1. Questions of a pre-experimental nature.
2. Questions which go beyond the first case.
3. "How" questions generally.
4. Questions calling for correlation or combination of multiple observations, data, evidences to give a response.
5. Explanation questions.

Examples:

1. Explain to me how that works.
2. How is that working?
3. Could you go on -- extension: is there any more to that?
4. Is there anything that holds true for all of these?
5. How do you account for this?
6. How could this have happened?

C. Deduction

1. Formulation Hypotheses: Teacher seeks to elicit responses of formulating hypotheses regarding the causes of observed phenomena.

   Includes:

   1. All questions which ask only for the causes not effects.
   2. Questions which focus on the causes of an observed problem.
   3. "Why" questions seeking cause.

   Examples:

   1. What hypotheses could you make about the causes of that condition?
   2. Why did that happen? (What caused it to happen?)
   3. What do you think is the cause of that?

2. Operational Definitions: Teacher seeks a statement from students requiring a definition of something in terms of its observable characteristics in such a way that another person can identify its presence or absence. Students may be asked to define in terms of the actual techniques used to detect it.

   Includes:

   1. Questions requiring students to make responses in terms of structure and function.

   Examples:

   1. What is an operational definition of a pencil?
   2. Can you describe this state or condition in terms of the measurement used to detect it?
   3. How could you define this so that others could detect it?

3. Manipulation and Controlling Variables: Students are given conditions under which the observations were made and they are asked to make observations when the conditions are varied. Or they are asked to make and state observations if/when the conditions are varied. "Variable" = a quantity that may assume a succession of values or states not limited to, but including, numerical values.

Includes:

1. Prediction-type questions in which one or more items are changed or varied to obtain an outcome immediately before or after.

Examples:

1. Suppose I increase the temperature, what happens?
2. If you remove the claws from the tiger, what would happen?
3. What would happen if I make this number a 4?

4. Developing an Experiment: Teacher asks students to plan an experiment to test a hypothesis. This is a formal procedure.

Includes:

1. Questions which ask how one could test a formal hypothesis or test an idea.
2. Questions eliciting statements of problem, methods or procedures, materials to be used in a formal context.

Examples:

1. What kind of experiment would you set up to test that?
2. How could we test that?
3. How could we determine whether that's true or not?

5. Interpreting Data: Teacher asks students to interpret experimental data and to draw conclusions from it. (The experiment has already been set up or completed in a laboratory or theoretical context.) Interpreting data questions must be the result of an experiment performed or described, or data in numerical, chart, or graph form from an experiment.

Includes:

1. Questions calling for conclusions.

Examples:

1. How would you interpret this chart or graph?
2. How would you interpret the results of that experiment?
3. What would be your conclusion?

II. NONINQUIRING QUESTIONS

The teacher asks questions of students which could be categorized as requiring statements of facts or relating facts. The questioning behavior is characterized by questions which seek answers that are previously known to the student.

Includes:

1. Questions answerable by "Yes" or "No."
2. Cue rhetorical words such as "Huh?," "O.K.?," "All right?," "Right?"
3. Self-reference questions. If there is problem solving involved on the spot, it's inquiry; if not, if response will be based on past material or self, personal reference, it is noninquiry.
4. Simple questions not requiring elaboration. Look for name of person, place, thing.
5. Teacher repeating a student response as a question.
6. Teacher asking student to repeat or clarify.
7. Teacher asking if students have questions.
8. Identification of labels which does not require abstracting physical characteristics, past or present, through the use of senses or thought.
9. Questions eliciting statements of observation made in past. (Prior to beginning of lesson.) If information is not discussed as part of the lesson, it is considered previous knowledge.
10. Questions asking for information, expository information.
11. If teacher asks student for answer student has already worked out in homework or independent class work.

Examples:

1. How did you personally feel about that?
2. How do you define "hypothesis"? When did Columbus discover America?
3. Where did it happen? What did the man say? What's the title?
4. What did we decide yesterday?
5. What did you say? Would you repeat that? Do you have questions?

III. UNCLASSIFIABLE

Incomplete questions, unintelligible questions, unclassifiable questions will be placed in this category.

Includes:

1. Questions with blanks or omission causing the question to be unintelligible.